**Journal Report 11**

**11/3/23**

I worked on preprocessing my dataset into the “Setpoint” and “Sensor A” attributes that I mentioned last time and was able to create a standardized format into a .csv file. Afterwards, I got back to work on implementing the GAN solution. I had some trouble importing the different libraries again since I was using Kaggle, so I spent the last 20 minutes of class time troubleshooting this issue. I wasn’t able to solve this problem through the same fix as last time, which involved creating a new anaconda environment and downloading the old version of the library. I am still thinking of different solutions, but I am leaning towards either getting a new implementation for GAN that doesn’t use the same library, trying to get Anaconda on a Kaggle notebook, or switching over to Jupyter Notebook. My caveat with Jupyter Notebook is that my computer only has 8 GB of ram so training the model would take a long time.

**11/7/23 and 11/8/23**

I spent the majority of my long weekend and Wednesday working on my application for the Regeneron Science Talent Search as it was due on Wednesday afternoon. As a result, I worked on documenting all my work and creating it into a research paper. This amounted to about 27 hours of work so I was not able to do much in terms of progress for the project.

For the competition, I created a 15 page research paper and had to write numerous 200 word prompts for both the data preparation and methodology of my research project. While I was writing the paper, I noticed that my previous preprocessing pipeline may promote noise. Specifically, I chose to apply the Savitsky-Golay filter after upsampling my data. It might be better to apply the filter before upsampling my data in order to preserve the long-term trends rather than every little peak as shown by Figure 2. Although the filter is smoothing over the noise, the time series itself after upsampling is over a 4 hour timespan, whereas the original dataset lasted for approximately 20 hours. Smoothing over the voltage values is making the line itself less noisy, but there is still a lot of noise on the scale of 20 hours. Values like the voltage could be indirectly affecting the performance of the model and the forecasting of other variables. Specifically, values like the “Set B field” and “B field” should not be as noisy as they are relatively smooth lines like the “Setpoint” and “Sensor A” values in the CCR dataset. I’m going to keep this in mind and try applying my filter with different parameter values by next Wednesday .



